

# Combining independent, arbitrarily weighted $P$ -values: a new solution to an old problem using a novel expansion with controllable accuracy

Gelio Alves

*National Institutes of Health, Bethesda, USA*

and

Yi-Kuo Yu<sup>†</sup>

*National Institutes of Health, Bethesda, USA*

**Summary.** Good's formula and Fisher's method are frequently used for combining independent  $P$ -values. Interestingly, the equivalent of Good's formula already emerged in 1910 and mathematical expressions relevant to even more general situations have been repeatedly derived, albeit in different context. We provide here a novel derivation and show how the analytic formula obtained reduces to the two aforementioned ones as special cases. The main novelty of this paper, however, is the explicit treatment of nearly degenerate weights, which are known to cause numerical instabilities. We derive a controlled expansion, in powers of differences in inverse weights, that provides both accurate statistics and stable numerics.

**Keywords:** combine  $P$ -values; nearly degenerate weights; exponential variable; gamma variable; gamma distribution; Erlang distribution

## 1. Introduction

The question of how to obtain an overall significance level for the results of *independent* runs of studies has been investigated since the 1930s (Tippett, 1931; Fisher, 1932; Pearson, 1933, 1938). In fact, forming a single statistical significance out of multiple independent tests has been an important subject of study in numerous area of scientific disciplines, including social psychology (Stouffer et al., 1949; Mosteller and Bush, 1954), medical research (Olkin, 1995), genetics (Loesgen et al., 2001), proteomics (Alves et al., 2008), genomics (Hess and Iyer, 2007), bioinformatics (Bailey and Gribskov, 1998; Yu et al., 2006) and others.

Frequently used methods for combining independent  $P$ -values fall into numeric and analytic categories. This classification is not totally precise since method such as Fisher's started out with the necessity of inverting the  $\chi^2$  cumulative distribution function and thus seemed like a numerical approach (Pearson, 1938). The method mentioned in (Bahucha-Reid, 1960), although not in the context of combining  $P$ -values, brought out an analytical expression for combined  $P$ -value using Fisher's method, thus effectively brought Fisher's method into analytic category. In the context of combining  $P$ -values, by mapping to a known result

<sup>†</sup>*Address for correspondence:* National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA.

E-mail: [yyu@ncbi.nlm.nih.gov](mailto:yyu@ncbi.nlm.nih.gov)

by Feller (Feller, 1966), Bailey and Gribskov (1998) also provided an analytic formula for Fisher's combined  $P$ -value. Numerical approaches typically involve inverting cumulative distribution functions. For example, Stouffer's  $z$ -methods (Stouffer et al., 1949), whether unweighted (Whitlock, 2005) or weighted (Liptak, 1958; Koziol and Tuckwell, 1994), require inverting the error function. Lancaster's generalization (Lancaster, 1961; Koziol, 1996) of Fisher's formalism also requires inverting gamma distribution function to incorporate unequal weighting for  $P$ -values combined.

In this paper, we focus on analytic methods only. Two existing analytic approaches, Fisher's (Fisher, 1932; Bahrucha-Reid, 1960; Bailey and Gribskov, 1998; Alves et al., 2008) and Good's (Good, 1955; Likes, 1967), are frequently employed. Fisher's method combines  $L$  independent tail-area probabilities democratically to form a single significance assignment while Good's formula weights each tail-area probability differently to form a different single significance assignment. Since Good's expression involves, in the denominator, pairwise differences between weights, he cautiously remarked that the expression may become ill-conditioned and thus the calculations should be done by holding more decimal places when weights of similar magnitudes exist. This statement has been paraphrased by many authors (Bhoj, 1992; Olkin and Saner, 2001; Hou, 2005).

In addition to the cases considered by Fisher and Good, it is foreseeable that one may wish to categorize obtained independent  $P$ -values into different groups so that one would like to weight  $P$ -values within the same group democratically and weight different group differently. We will call this scenario the *general case* (GC). The GC naturally occurs since one may wish to categorize data obtained from the same type of experimental instruments into the same group, and data collected from different instrument types may justify the use of different weights. When there is only one instrument type, the GC reduces to Fisher's consideration. When there exist no replicates within each instrument types, the GC coincides with the consideration of Good. In principle, the weighted version of the Stouffer's method can also be used for this purpose. Since the main scope here is to pursue analytic results, we won't delve into methods in numerical category.

It is important, however, to point out that the mathematical problem of combining  $P$ -values is also related to other areas of research. For example, the equivalent of Good's formula had already emerged in 1910 in the context of sequential radioactive decay (Bateman, 1910), while analytical expression for Fisher's combined  $P$ -value emerged in 1960 as a special case of the former when all the decay constants are identical (Bahrucha-Reid, 1960). After Good's formula (Good, 1955), the same distribution function was rederived by McGill and Gibbon (1965) and later on by Likes (1967). As for the GC, Fisher's method included, the mathematical equivalents appear in different areas of studies mainly under the consideration of sum of exponential/gamma variables. The distribution function of linear combination of exponential/gamma variables are useful in various fields. When limited to exponential variables, it results in the Erlang distribution that is often encountered in queuing theory (Morse, 1958). It is also connected to the renewal theory (Cox, 1962), time series problem (MacNeill, 1974), and can be applied to model reliability (Jasiulewicz and Kordecki, 2003). The intimate connections between these seemingly different problems are not obvious at first glance. Consequently, it is not surprising that the distribution function has been rediscovered/rederived many times and that some information about it has not been widely circulated. Our literature searches show that the first explicit result (without further derivatives involved) for the distribution function was obtained by Mathai (1983). Subsequently, motivated by different contexts, Harrison (1990), Amari and Mirsa (1997), and Jasiulewicz and Kordecki (2003) all rederived the

same distribution function.

Employing a complex variable integral formulation, we are able to provide a different derivation of the distribution function and become the first to make connection to the GC of combining  $P$ -values. Since both Fisher's and Good's considerations arise as special limiting cases of the GC, we also illustrate that our cumulative probability distribution for GC indeed reduces to the appropriate limiting formulas upon taking appropriate parameters. The main novelty of this paper, however, is the explicit treatment of cases where nearly degenerate weights exist. These cases are known for numerical instabilities, which motivated many authors to pursue uncontrolled approximations (Solomon and Stephens, 1977; Gabler, 1987; Bhoj, 1992; Hou, 2005). We have derived a controlled expansion, in power of differences in inverse weights, that provides both accurate statistics and stable numerics.

In the following section, we will first summarize Fisher's and Good's methods for combining  $P$ -values, followed by the mathematical definition of the GC. A section devoted to derivation of the probability distribution function and cumulative probability for the GC then follows. We then delve into the case of nearly degenerate weights and provides a formula with controllable accuracy for combining  $P$ -values. A few examples of using the main results are then provided. This paper concludes with future directions. A C++ program, which computes the combined  $P$ -values with equal numerical stability regardless of whether weights are (nearly) degenerate or not, is available upon request from the authors.

## 2. Summary of Fisher's and Good's methods for combining $P$ -values

Imagine that one wishes to combine  $L$  independent  $P$ -values  $p_1, p_2, \dots, p_L$ , each of which is drawn from an uniform, independent distribution over  $(0, 1]$ . For later convenience, let us define

$$\tau_F \equiv p_1 \cdot p_2 \cdots p_L, \quad (1)$$

$$\tau_G \equiv p_1^{w_1} \cdot p_2^{w_2} \cdots p_L^{w_L}. \quad (2)$$

To form a unified significance, Fisher and Good considered respectively the stochastic quantities  $Q_F$  and  $Q_G$ , defined by

$$Q_F \equiv x_1 \cdot x_2 \cdots x_L, \quad (3)$$

$$Q_G \equiv x_1^{w_1} \cdot x_2^{w_2} \cdots x_L^{w_L}, \quad (4)$$

where each  $x_i$  represents a random variable drawn from an uniform, independent distribution over  $(0, 1]$ . The following probabilities

$$\text{Prob}(Q_F \leq \tau_F) = \tau_F \sum_{l=0}^{L-1} \frac{[\ln(1/\tau_F)]^l}{l!} \quad (5)$$

$$\text{Prob}(Q_G \leq \tau_G) = \sum_{l=1}^L \Lambda_l \tau_G^{1/w_l} \quad (6)$$

provide the unified statistical significances, corresponding respectively to Fisher's and Good's considerations, from combining  $L$  independent  $P$ -values. In eq. (6), the prefactor  $\Lambda_l$  is given by

$$\Lambda_l = \frac{w_l^{L-1}}{\prod_{k \neq l} (w_l - w_k)}. \quad (7)$$

Apparently,  $\Lambda_l$  is ill-defined when the weight  $w_l$  coincides with or is numerically close to any other weights  $w_k$ . Although Fisher did not derive (5), from this point on, we shall refer to (5) as Fisher's formula and (6) as the Good's formula.

### 3. General case including Fisher's and Good's formulas

Let us divide the  $L$   $P$ -values into  $m$  groups with  $1 \leq m \leq L$ . Within each group  $k$ , we weight the  $n_k$   $P$ -values within equally; while  $P$ -values in different groups are weighted differently. Therefore, when  $m = L$  and  $n_k = 1 \ \forall k$ , we have the Good's case and when  $m = 1$  and  $n_1 = L$ , we reach Fisher case. We will therefore define the following quantities of interest

$$\tau \equiv \prod_{k=1}^m \left[ \prod_{j=1}^{n_k} p_{k;j} \right]^{w_k} \quad (8)$$

$$Q \equiv \prod_{k=1}^m \left[ \prod_{j=1}^{n_k} x_{k;j} \right]^{w_k} \quad (9)$$

where each  $x_{k;j}$  represents again a random variable drawn from an uniform, independent distribution over  $(0, 1]$ . The quantity of interest  $\text{Prob}(Q \leq \tau)$ , if obtained, should cover both results of Fisher's and Good's as the limiting cases. In the next section, we will derive an exact expression for  $\text{Prob}(Q \leq \tau)$  and describe how to recover the results of Fisher's and Good's.

### 4. Derivation of $\text{Prob}(Q \leq \tau)$

Let  $F(\tau) \equiv \text{Prob}(Q \leq \tau)$ , we then may write

$$F(\tau) = \int_0^1 \cdots \int_0^1 \theta \left( \tau - \prod_{k=1}^m \left[ \prod_{i=1}^{n_k} x_{k;i} \right]^{w_k} \right) \prod_{k=1}^m \prod_{j=1}^{n_k} dx_{k;j} , \quad (10)$$

where the heaviside step function  $\theta(x)$  takes value 1 when  $x > 0$  and value 0 when  $x < 0$ . Upon taking a derivative with respect to  $\tau$ , we obtain

$$f(\tau) \equiv \frac{dF(\tau)}{d\tau} = \int_0^1 \cdots \int_0^1 \delta \left( \tau - \prod_{k=1}^m \left[ \prod_{j=1}^{n_k} x_{k;j} \right]^{w_k} \right) \prod_{k=1}^m \prod_{j=1}^{n_k} dx_{k;j} , \quad (11)$$

where  $\delta(x)$  represents Dirac's delta function that takes zero value everywhere except at  $x = 0$  and that  $\forall a > 0$ ,  $\int_{-a}^a \delta(x) dx = 1$ .

To proceed, let us make the following change of variables

$$\begin{aligned} \tau &= e^{-t} \\ x_{k;j} &= e^{-u_{k;j}} \end{aligned}$$

and remember that if  $y_0$  is the only root of  $f$  ( $f(y_0) = 0$ )

$$\delta(f(y)) = \frac{\delta(y - y_0)}{|f'(y_0)|} ,$$

we may then rewrite (11) as

$$f(\tau) = \int_0^\infty \cdots \int_0^\infty e^t e^{-\sum_{k,j} u_{k;j}} \delta \left( t - \sum_{k=1}^m w_k \left[ \sum_{j=1}^{n_k} u_{k;j} \right] \right) \prod_{k=1}^m \prod_{j=1}^{n_k} du_{k;j} . \quad (12)$$

Note that the right hand side of (12), except for the additional factor  $e^t$ , is the probability density function of a weighted, linear sum of exponential variables.

By introducing the integral representation of the  $\delta$  function

$$\delta(t - c) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dq e^{-iq(t-c)} ,$$

we may re-express (12) as

$$\begin{aligned} f(\tau) &= \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{k=1}^m \left[ \int_0^\infty e^{-u} e^{iqw_k u} du \right]^{n_k} \\ &= \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{k=1}^m \left[ \frac{1}{1 - iqw_k} \right]^{n_k} \\ &= \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{l=1}^m \left( \frac{i}{w_l} \right)^{n_l} \prod_{k=1}^m \left[ \frac{1}{q + ir_k} \right]^{n_k} \\ &= \left[ \prod_{l=1}^m r_l^{n_l} \right] (i)^{\sum_{k=1}^m n_k} \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{k=1}^m \left[ \frac{1}{q + ir_k} \right]^{n_k} \end{aligned} \quad (13)$$

$$\equiv \left[ \prod_{l=1}^m r_l^{n_l} \right] \tilde{f}(\tau; n_1, n_2, \dots, n_m) , \quad (14)$$

where  $r_k \equiv 1/w_k$  is introduced for the ease of analytical manipulation and  $\tilde{f}$  is introduced for later convenience. Since all  $w_k > 0$ , implying that all  $r_k > 0$  and thus the poles of the integrand in (13) lie completely at the lower half of the  $q$ -plane. The integral of  $q$  may be extended to enclose the lower half  $q$ -plane to result in

$$\begin{aligned} f(\tau) &= e^t \left[ \prod_{l=1}^m (i r_l)^{n_l} \right] \left( \frac{-2\pi i}{2\pi} \right) \sum_{k=1}^m \frac{(\partial/\partial q)^{n_k-1}}{(n_k-1)!} \left[ e^{-itq} \prod_{j=1, j \neq k}^m \left( \frac{1}{q + ir_j} \right)^{n_j} \right]_{q=-ir_k} \\ &= e^t \left[ \prod_{l=1}^m (i r_l)^{n_l} \right] \sum_{k=1}^m \left\{ (-i) \sum_{\substack{g_1, g_2, \dots, g_m=0 \\ \sum g_i = n_k-1}} \frac{(-1)^{n_k-1} (it)^{g_k}}{g_k! e^{r_k t}} \prod_{j=1, j \neq k}^m \frac{(n_j-1+g_j)!}{(n_j-1)! g_j!} \left( \frac{-i}{r_j - r_k} \right)^{n_j+g_j} \right\} \\ &= e^t \left[ \prod_{l=1}^m r_l^{n_l} \right] \sum_{k=1}^m \left\{ \sum_{\substack{g_1, g_2, \dots, g_m=0 \\ \sum g_i = n_k-1}} \frac{(t)^{g_k}}{g_k!} e^{-r_k t} \prod_{j=1, j \neq k}^m \frac{(n_j-1+g_j)!}{(n_j-1)! g_j!} \frac{(-1)^{g_j}}{(r_j - r_k)^{n_j+g_j}} \right\} . \end{aligned} \quad (15)$$

Comparing eq. (15) with eq. (12), we see that the right hand side of (15) is composed of the product of the factor  $e^t$  and the probability density function of a weighted sum

of exponential variables. In fact, the explicit expression of latter, in addition to the new derivation presented here, was derived much earlier (Mathai, 1983) under different context and was rediscovered/rederived multiple times (Harrison, 1990; Amari and Mirsa, 1997; Jasiulewicz and Kordecki, 2003) by different means. Its connection to combining  $P$ -values, however, was never made explicit till now.

From (10), we know that  $F(\tau = 0) = 0$ , implying that

$$\begin{aligned}
F(\tau) &= \int_0^\tau f(\tau') d\tau' = \int_t^\infty f(e^{-t'}) e^{-t'} dt' \\
&= \left[ \prod_{l=1}^m r_l^{n_l} \right] \sum_{k=1}^m \sum_{\substack{g_1, g_2, \dots, g_m=0 \\ \sum g_i = n_k - 1}} \left( \prod_{j=1, j \neq k}^m \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \frac{(-1)^{g_j}}{(r_j - r_k)^{n_j + g_j}} \right) \int_t^\infty \frac{(t')^{g_k}}{g_k!} e^{-r_k t'} dt' \\
&= \left[ \prod_{l=1}^m r_l^{n_l} \right] \sum_{k=1}^m \sum_{\substack{g_1, g_2, \dots, g_m=0 \\ \sum g_i = n_k - 1}} \left( \prod_{j=1, j \neq k}^m \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \frac{(-1)^{g_j}}{(r_j - r_k)^{n_j + g_j}} \right) \left( \sum_{l=0}^{g_k} \frac{t^{g_k - l}}{r_k^{l+1} (g_k - l)!} e^{-r_k t} \right) \\
&= \sum_{k=1}^m \sum_{\substack{g_1, g_2, \dots, g_m=0 \\ \sum g_i = n_k - 1}} \left( \prod_{j=1, j \neq k}^m \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \frac{(-r_k)^{g_j} r_j^{n_j}}{(r_j - r_k)^{n_j + g_j}} \right) \left( \sum_{l=0}^{g_k} \frac{(r_k t)^{g_k - l}}{(g_k - l)!} e^{-r_k t} \right) \\
&= \sum_{k=1}^m \sum_{g_k=0}^{n_k-1} \sum_{\substack{g_i \neq k=0 \\ \sum_i g_i = n_k - 1}}^{n_k-1-g_k} \left( \prod_{j=1, j \neq k}^m \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \frac{(-r_k)^{g_j} r_j^{n_j}}{(r_j - r_k)^{n_j + g_j}} \right) H(r_k t, g_k), \tag{16}
\end{aligned}$$

where the function  $H$  is defined as

$$H(x, n) \equiv e^{-x} \sum_{k=0}^n \frac{x^k}{k!}. \tag{17}$$

Eq. (16) represents the most general formula that interpolates the scenarios considered by Fisher and Good.

Let us take the limiting case from (16). For Fisher's formula, one weights every  $P$ -value equally, and thus correspond to  $m = 1$  and  $n_1 = L$ . The constraint in the sum of (16) forces  $g_1 = n_1 - 1 = L - 1$ . Consequently, we have (by calling  $r_1$  by  $r$  for simplicity)

$$\text{Prob}(Q_F \leq \tau_F) = H(rt, L - 1) = e^{-rt} \sum_{l=0}^{L-1} \frac{(rt)^l}{l!} \tag{18}$$

Notice that regardless whatever the weight  $w$  one assigns to all the  $P$ -values, the final answer is independent of the weight. This is because  $t = -\ln \tau = -w \ln \tau_F = (-\ln \tau_F)/r$  and therefore  $rt = \ln(1/\tau_F)$ . This results in

$$\text{Prob}(Q_F \leq \tau_F) = \tau_F \sum_{l=0}^{L-1} \frac{[\ln(1/\tau_F)]^l}{l!}, \tag{19}$$

exactly what one anticipates from (5). To obtain the results of Good, one simply makes  $m = L$  and  $n_k = 1 \forall k$ , implying all  $g_i = 0$ . In this case, (16) becomes (with  $r_l = 1/w_l$ ,

$e^{-t} = \tau_G$  and  $H(x, 0) = 1 \forall x$ )

$$\text{Prob}(Q_G \leq \tau_G) = \sum_{l=1}^L \left( \prod_{k \neq l} \frac{r_k}{r_k - r_l} \right) \tau_G^{1/w_l} = \sum_{l=1}^L \Lambda_l \tau_G^{1/w_l}, \quad (20)$$

reproducing exactly (6)

One may also re-express eq. (16) in a slightly different form

$$\begin{aligned} F(\tau) &= \left[ \prod_{l=1}^m r_l^{n_l} \right] \sum_{k=1}^m \sum_{g_k=0}^{n_k-1} \frac{1}{r_k^{g_k+1}} H(r_k t, g_k) \times \\ &\quad \times \sum_{\substack{g_i \neq k=0 \\ \sum_i g_i = n_k - 1}}^{n_k-1-g_k} \left( \prod_{j=1, j \neq k}^m \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \frac{(-1)^{g_j}}{(r_j - r_k)^{n_j + g_j}} \right) \\ &\equiv \left[ \prod_{l=1}^m r_l^{n_l} \right] \tilde{F}(\tau; n_1, n_2, \dots, n_m). \end{aligned} \quad (21)$$

Note that in the expression (21), we have isolated an overall multiplying factor and keeps explicit the  $n_{1 \leq k \leq m}$  dependence for later convenience.

## 5. Cases of nearly degenerate weights

In our derivation of (19), it is explicitly shown in section 4 that the final  $P$ -value obtained is independent of the weight  $w$  that was used to assign to all the individual  $P$ -values,  $p_1, p_2, \dots, p_L$ . It is thus natural to ask, if one starts by weighting each  $P$ -value differently, upon making the weights closer to one another, will one recover Fisher's formula (5) from Good's formula (6)? By continuity, the answer is expected to be affirmative. More generally, one would like to have a formal protocol to compute the combined  $P$ -value when the weights may be categorized into several subsets, within each subset the weights are almost degenerate.

In this section, we first illustrate the transition from Good's formula to Fisher's formula by combining two almost degenerate  $P$ -values. We will then provide a general protocol to explicitly, when there exist nearly degenerate weights, deal with the possible numerical instability that was first cautioned by Good (1955) and subsequently by many authors (Bhoj, 1992; Olkin and Saner, 2001; Hou, 2005).

Let us consider combining  $p_1$  and  $p_2$  with weights  $w_1$  and  $w_2$  using Good's formula. One has

$$\text{Prob}(Q_G \leq \tau_G) = \frac{1}{w_1 - w_2} \left[ w_1 p_1 p_2^{\frac{w_2}{w_1}} - w_2 p_1^{\frac{w_1}{w_2}} p_2 \right]. \quad (22)$$

Without loss of generality, one assumes  $w_1 > w_2$  and hence writes  $w_1/w_2 = 1 + \epsilon$  with  $\epsilon > 0$ . We are interested in the case when the weights get close to each other, or when  $\epsilon \rightarrow 0$ . We now rewrite eq. (22) as

$$\text{Prob}(Q_G \leq \tau_G) = \frac{w_2}{w_1 - w_2} \left[ \frac{w_1}{w_2} p_1 p_2^{\frac{w_2}{w_1}} - p_1^{\frac{w_1}{w_2}} p_2 \right] = \frac{1}{\epsilon} \left[ (1 + \epsilon) p_1 p_2^{\frac{1}{1+\epsilon}} - p_1^{1+\epsilon} p_2 \right]. \quad (23)$$

In the limit of small  $\epsilon$ , we may rewrite (23) as

$$\begin{aligned}
\text{Prob}(Q_G \leq \tau_G) &= \frac{p_1 p_2}{\epsilon} \left[ (1 + \epsilon) p_2^{-\frac{\epsilon}{1+\epsilon}} - p_1^\epsilon \right] = \frac{p_1 p_2}{\epsilon} \left[ (1 + \epsilon) e^{-\frac{\epsilon}{1+\epsilon} \ln p_2} - e^{\epsilon \ln p_1} \right] \\
&= \frac{p_1 p_2}{\epsilon} \left[ \epsilon - \epsilon (\ln p_2 + \ln p_1) + \mathcal{O}(\epsilon^2) \right] \\
&= p_1 p_2 [1 - \ln(p_1 p_2) + \mathcal{O}(\epsilon)] \\
&\xrightarrow{\epsilon \rightarrow 0} p_1 p_2 [1 - \ln(p_1 p_2)] = \text{Prob}(Q_F \leq \tau_F) .
\end{aligned} \tag{24}$$

Note that when the small weight difference  $w_1 - w_2$  is near the machine precision of a digital computer, using formula (6) directly will inevitably introduce numerical instability caused by rounding errors.

To construct a general protocol to deal with nearly degenerate weights, one first observes from eqs. (13-21) that it is the inverse of weights  $r_k \equiv 1/w_k$  that permeates the derivation of the unified  $P$ -value. The closeness between weights is thus naturally defined by closeness in the inverse weights. As shown in eqs. (2) and (6), the combined  $P$ -value by Good's formula is independent of the absolute size of the weights but only on the relative weights. Making the observation that  $r_k t$  in eq. (16) only depend on the ratios  $r_{j \neq k}/r_k$ , one also sees explicitly that the most general combined  $P$ -values (see (16)) only depend on the relative weights as well. We are thus free to choose any scale we wish. For simplicity, we normalize the inverse weight associated with each method by demanding the sum of inverse weights equal the total number of methods

$$\sum_{j=1}^M r_j = \sum_{j=1}^M 1 = M , \tag{25}$$

where  $1/r_j$  represents the weight associated method  $j$  and  $M$  represents the total number of  $P$ -values (or methods) to be combined. For the GC described in section 4,  $M = \sum_{k=1}^m n_k$ . This normalization choice makes the average inverse weight of participating methods be 1.

The next step is to determine, for a given list of inverse weights and the radius of clustering, the number of clusters needed. This task may be achieved in a hierarchical manner. After normalizing the inverse weights  $r_k$  using eq. (25), one may sort the inverse weights in either ascending or descending order. For a given radius  $\eta > 0$ , one starts to seek the pair of inverse weights that are closest but not identical, and check if it is smaller than the radius  $\eta$ . If yes, one will merge that pair of inverse weights by using their average, weighted by number of occurrences, as the new center and continue the process till every inverse weights in the list is separated by a distance farther than  $\eta$ . We use an example of  $M = 8$  to illustrate the idea. Let the normalized inverse weights  $\{r_j\}_{j=1}^8$  be

$$0.50, 0.70(2), 0.71, 0.74, 1.03, 1.80, 1.82$$

where the number 2 inside the pair of parentheses after 0.70 simply indicates that there are two identical inverse weights 0.70 to start with. Assume that one chooses the radius of cluster  $\eta$  to be 0.005, since every adjacent inverse weights are separated by more than 0.005, no further clustering procedures is needed and one ends up having seven effective clusters: one cluster with two identical inverse weights 0.70, and the rest of six clusters are all singletons. This corresponds to  $m = 7$ ,  $n_1 = 1$ ,  $n_2 = 2$ ,  $n_3 = n_4 = \dots = n_7 = 1$ .



Suppose one chooses the clustering radius  $\eta$  to be 0.05. In the first step, we identify that 0.70 and 0.71 are the closet pair of inverse weights. The weighted average between them is

$$\frac{2 \cdot 0.70 + 0.71}{3} = \frac{2.11}{3} = 0.70\bar{3}.$$

The list of inverse weights then appears as

$$0.50, 0.70\bar{3}(3), 0.74, 1.03, 1.80, 1.82.$$

The closest pair of inverse weights is now between 1.80 and 1.82, and upon merging them we have the list now appears as

$$0.50, 0.70\bar{3}(3), 0.74, 1.03, 1.81(2).$$

Next pair of closest inverse weights is then 0.70 $\bar{3}$  and 0.74. The weighted average leads to  $(2.11 + 0.74)/4 = 0.7125$ . After this step, the list of inverse weights appears as

$$0.50, 0.7125(4), 1.03, 1.81(2),$$

indicating that we have  $m = 4$  ( four clusters), with number of members being  $n_1 = 1$ ,  $n_2 = 4$ ,  $n_3 = 1$  and  $n_4 = 2$ . The centers of the four clusters are specified by the average inverse weights: 0.50, 0.7125, 1.03, 1.81. The distance between any two of the average inverse weights is now larger than 0.05.

This is a good place for us to introduce some notation. We shall denote by  $r_k + \eta_{k;j}$  the  $j$ th inverse weights of cluster  $k$ , whose averaged inverse weight is  $r_k$ . With this definition, for the example above, we have  $\eta_{1;1} = 0$ ,  $\eta_{2;1} = \eta_{2;2} = -0.0125$ ,  $\eta_{2;3} = -0.0025$ ,  $\eta_{2;4} = 0.0275$ ,  $\eta_{3;1} = 0$ ,  $\eta_{4;1} = -0.01$ , and  $\eta_{4;2} = 0.01$ .

Using the hierarchical protocol mentioned above, the number of clusters  $m$  and the numbers of members  $n_k$  associated with cluster  $k$  are all obtained along with  $\{\eta_{k;j}\}$ . Following the derivation in section 4, we obtain a probability density function very similar to (13)

$$f(\tau) = \left[ \prod_{l=1}^m \prod_{j=1}^{n_l} (r_l + \eta_{l;j}) \right] (i)^{\sum_{k=1}^m n_k} \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{k=1}^m \left[ \prod_{j=1}^{n_k} \frac{1}{q + i(r_k + \eta_{k;j})} \right]. \quad (26)$$

From section 4, we see that the ill-conditioned situations emerge when some weights are nearly degenerate and the source of difference in inverse weights comes from obtaining  $\tilde{F}(\tau; n_1, n_2, \dots, n_m)$  in (21) from  $\tilde{f}(\tau; n_1, n_2, \dots, n_m)$  in (14). Therefore, one may leave the prefactor  $\left[ \prod_{l=1}^m \prod_{j=1}^{n_l} (r_l + \eta_{l;j}) \right]$  untouched and focus on the rest of the right hand side of eq. (26). To proceed, we write

$$\begin{aligned} \frac{1}{q + i(r_k + \eta_{k;j})} &= \frac{1}{q + ir_k} \left( 1 + \frac{i \eta_{k;j}}{q + ir_k} \right)^{-1} = \frac{1}{q + ir_k} e^{-\ln\left(1 + \frac{i \eta_{k;j}}{q + ir_k}\right)} \\ &= \frac{1}{q + ir_k} \exp \left[ \sum_{g=1}^{\infty} \frac{1}{g} \left( \frac{-i \eta_{k;j}}{q + ir_k} \right)^g \right]. \end{aligned}$$

Consequently, we may write

$$\prod_{j=1}^{n_k} \frac{1}{q + i(r_k + \eta_{k;j})} = \frac{1}{(q + ir_k)^{n_k}} \exp \left[ \sum_{g=1}^{\infty} \frac{Y_{k;g}(i)^g}{(q + ir_k)^g} \right] \quad (27)$$

where

$$Y_{k;g} \equiv \sum_{j=1}^{n_k} \frac{(-\eta_{k;j})^g}{g}. \quad (28)$$

The product in eq. (26) may now be formally written as

$$\prod_{k=1}^m \left[ \prod_{j=1}^{n_k} \frac{1}{q + i(r_k + \eta_{k;j})} \right] = \left[ \prod_{k=1}^m \frac{1}{(q + ir_k)^{n_k}} \right] \exp \left[ \sum_{g=1}^{\infty} (i)^g \sum_{k=1}^m \frac{Y_{k;g}}{(q + ir_k)^g} \right]. \quad (29)$$

We now note a simplification by choosing  $r_k$  to be the average inverse weight of the  $k$ th cluster. In this case, we have  $\sum_{j=1}^{n_k} \eta_{k;j} = 0 \forall k$ . That is,  $Y_{k;1} = 0$  always. This allows us to write eq. (29) as

$$\prod_{k=1}^m \left[ \prod_{j=1}^{n_k} \frac{1}{q + i(r_k + \eta_{k;j})} \right] = \left[ \prod_{k=1}^m \frac{1}{(q + ir_k)^{n_k}} \right] \exp \left[ \sum_{g=2}^{\infty} (i)^g \sum_{k=1}^m \frac{Y_{k;g}}{(q + ir_k)^g} \right]. \quad (30)$$

The key idea here is to Taylor expand the exponential and collect terms of equal number of  $1/(q+ir)$ . Evidently, the first correction term starts with  $1/(q+ir)^2$ . Furthermore, before the  $1/(q+ir)^4$  order, there is no mixing between different clusters. Below, we rewrite eq. (26) to include the first few orders of correction terms

$$\begin{aligned} \frac{f(\tau)}{\prod_{l=1}^m \prod_{j=1}^{n_l} (r_l + \eta_{l;j})} &= (i)^{\sum_{k=1}^m n_k} \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \frac{\exp \left[ \sum_{g=2}^{\infty} (i)^g \sum_{k=1}^m \frac{Y_{k;g}}{(q + ir_k)^g} \right]}{\prod_{k=1}^m (q + ir_k)^{n_k}} \\ &= \tilde{f}(\tau; \{n_l\}_{l=1}^m) + \sum_{k=1}^m Y_{k;2} \tilde{f}(\tau; \{n_{l \neq k}, n_k + 2\}) \\ &\quad + \sum_{k=1}^m Y_{k;3} \tilde{f}(\tau; \{n_{l \neq k}, n_k + 3\}) + \sum_{k=1}^m \left( Y_{k;4} + \frac{(Y_{k;2})^2}{2!} \right) \tilde{f}(\tau; \{n_{l \neq k}, n_k + 4\}) \\ &\quad + \frac{1}{2!} \sum_{\substack{k,k'=1 \\ k \neq k'}}^m Y_{k;2} Y_{k';2} \tilde{f}(\tau; \{n_{l \neq k,k'}, n_k + 2, n_{k'} + 2\}) + \mathcal{O}(\eta^5). \end{aligned} \quad (31)$$

This immediately leads to

$$\begin{aligned} \frac{F(\tau)}{\prod_{l=1}^m \prod_{j=1}^{n_l} (r_l + \eta_{l;j})} &= \tilde{F}(\tau; \{n_l\}_{l=1}^m) + \sum_{k=1}^m Y_{k;2} \tilde{F}(\tau; \{n_{l \neq k}, n_k + 2\}) \\ &\quad + \sum_{k=1}^m Y_{k;3} \tilde{F}(\tau; \{n_{l \neq k}, n_k + 3\}) + \sum_{k=1}^m \left( Y_{k;4} + \frac{(Y_{k;2})^2}{2!} \right) \tilde{F}(\tau; \{n_{l \neq k}, n_k + 4\}) \\ &\quad + \frac{1}{2!} \sum_{\substack{k,k'=1 \\ k \neq k'}}^m Y_{k;2} Y_{k';2} \tilde{F}(\tau; \{n_{l \neq k,k'}, n_k + 2, n_{k'} + 2\}) + \mathcal{O}(\eta^5). \end{aligned} \quad (32)$$

Note that when the cluster radius  $\eta$  is chosen to be zero, the only clusters are from sets of *identical* weights, and all  $\eta_{k;j}$  must be zero. In this case, only the first term on the

right hand side of (32) exists and the result derived in section 4 is recovered exactly. Since all  $\tilde{F}$  are finite positive quantities, the errors resulting from truncating the expression in eq. (32) at certain order of  $\eta$  can be easily bounded. Therefore, any desired precision may be obtained via including the corresponding number of higher order terms. As the main result of the current paper, our expansion provides a systematic, numerically stable method to achieve desired accuracy in computing combined  $P$ -values.

## 6. Examples

**Example (a):** This example, assuming  $m = 4$ , demonstrates how to compute the  $\tilde{F}(\tau; \{n_l\})$  function present in eq. (21). Let  $r_k$  be the inverse weights associated with cluster  $k$ . When combining multiple  $P$ -values with a prescribed clustering radius on the inverse weights, see (25) and the previously described clustering procedure, the  $r_k$ s and the deviations  $\eta_{k;j}$  are obtained once and for all. The  $\eta_{k;j}$ , through eq. (28), constitute the key expansion parameters  $Y_{k;g}$  that yield, upon multiplying by  $\tilde{F}(\tau; n_l)$  with different  $\{n_l\}$ , the higher order terms in our key result (32). Note that in eq. (32), in the zeroth order term, the argument  $n_l$  of  $\tilde{F}$  represents the number of members associated with cluster  $l$ . However, for higher order correction terms, the  $n_l$ s entering  $\tilde{F}$  no longer carry the same meaning. Therefore, in the example shown here, one does not assume that  $n_j$  is the number of methods associated with cluster  $k$ . We now illustrate how to open up the sum in eq. (21). The constraint  $\sum_i g_i = n_k - 1$  implies that one only has  $m - 1$  independent  $g_i$ s. Once  $m - 1$   $g_i$ s are specified, the remaining one is also determined. To simplify the exposition, let us introduce the following notation

$$\alpha(g_j; j, k) \equiv \frac{(n_j - 1 + g_j)!}{(n_j - 1)!g_j!} \frac{(-1)^{g_j}}{(r_j - r_k)^{n_j + g_j}}.$$

One then expand the sum in (21) as

$$\begin{aligned} \tilde{F}(\tau) = & \sum_{g_1=0}^{n_1-1} \frac{H(r_1 t, g_1)}{r_1^{g_1+1}} \sum_{g_2=0,}^{n_1-1-g_1} \alpha(g_2; 2, 1) \sum_{g_3=0}^{n_1-1-g_1-g_2} \alpha(g_3; 3, 1) \alpha(g_4; 4, 1) \\ & + \sum_{g_2=0}^{n_2-1} \frac{H(r_2 t, g_2)}{r_2^{g_2+1}} \sum_{g_1=0}^{n_2-1-g_2} \alpha(g_1; 1, 2) \sum_{g_3=0}^{n_2-1-g_2-g_1} \alpha(g_3; 3, 2) \alpha(g_4; 4, 2) \\ & + \sum_{g_3=0}^{n_3-1} \frac{H(r_3 t, g_3)}{r_3^{g_3+1}} \sum_{g_1=0}^{n_3-1-g_3} \alpha(g_1; 1, 3) \sum_{g_2=0}^{n_3-1-g_3-g_1} \alpha(g_2; 2, 3) \alpha(g_4; 4, 3) \\ & + \sum_{g_4=0}^{n_4-1} \frac{H(r_4 t, g_4)}{r_4^{g_4+1}} \sum_{g_1=0}^{n_4-1-g_4} \alpha(g_1; 1, 4) \sum_{g_2=0}^{n_4-1-g_4-g_1} \alpha(g_2; 2, 4) \alpha(g_3; 3, 4). \end{aligned} \quad (33)$$

**Example (b):** This example illustrates the possibility of numerical instability associated with eqs. (6) and (21) when they are used to combine  $P$ -values with nearly equal weights. We also show how such instabilities are resolved by using eq. (32). Consider the case of combining five  $P$ -values,  $\{0.008000257, 0.008579261, 0.0008911761, 0.006967988, 0.004973110\}$ ,

weighted respectively by  $\{0.54531152, 0.54532057, 0.54531221, 0.54531399, 0.54531776\}$ . Using eq. (2), one obtains  $\tau_G = 4.30656196 \times 10^{-7}$ . The combined  $P$ -value is then obtained as the probability of attaining a random variable  $Q_G$ , defined in eq. (4), such that is less than or equal to  $\tau_G$ .

Combining  $P$ -values using eq. (6) gives

$$\begin{aligned} \text{Prob}(Q_G \leq \tau_G) &= 1923475672.53812003 + 134195847.49348195 \\ &\quad - 3271698577.16100168 + 1726093852.57087326 - 512066795.44147670 \\ &= -0.00000322. \end{aligned}$$

When one uses equation (21),  $\tau$  takes the value of  $\tau_G$  and the random variable  $Q$  is simply  $Q_G$ , and the combined  $P$ -value becomes

$$\begin{aligned} \text{Prob}(Q \leq \tau) &= 170090507.09336647 + 21761086.68190728 \\ &\quad - 972903041.25101399 + 941269625.31004059 - 512066795.44252247 \\ &= -0.00000006. \end{aligned}$$

Apparently, probability can't be negative and the negative values shown above illustrate how eqs. (6) and (21) may suffer from numerical instability when the weights are nearly degenerate. This numerical instability is removed by applying equation (32) which combines weighted  $P$ -values using a controlled expansion and yields, for this example,

$$\begin{aligned} \text{Prob}(Q \leq \tau) &= 5.379093 \times 10^{-8} + 1.407305 \times 10^{-16} \\ &\quad - 1.066323 \times 10^{-21} + 1.634917 \times 10^{-25} + \mathcal{O}(10^{-29}) \\ &= 5.37909 \times 10^{-8} \end{aligned}$$

**Example (c):** One natural question to ask is that how well does eq. (32) work when one chooses a larger clustering radius and group weights that are clearly distinguishable into a few clusters? To consider this case, let us use the five  $P$ -values from example (b) above but with weights chosen differently. Let us assume that the inverse weights ( $r_k \equiv 1/w_k$ ) associated with these five  $P$ -values are  $\{0.6, 0.65, 1.2, 1.25, 1.3\}$ . For this case,  $\tau = \tau_G = 1.935663 \times 10^{-13}$ . Combining  $P$ -value using formulas (6) yields

$$\begin{aligned} \text{Prob}(Q \leq \tau) &= 2.187324 \times 10^{-6} - 5.946040 \times 10^{-7} + 2.131226 \times 10^{-13} \\ &\quad - 8.011644 \times 10^{-14} + 7.639290 \times 10^{-15} \\ &= 1.59272 \times 10^{-6}, \end{aligned}$$

while combining  $P$ -values using (21) yields identical results

$$\begin{aligned} \text{Prob}(Q \leq \tau) &= 1.725699 \times 10^{-6} - 3.049251 \times 10^{-7} + 1.311524 \times 10^{-13} \\ &\quad - 6.162803 \times 10^{-14} + 7.639290 \times 10^{-15} \\ &= 1.59272 \times 10^{-6}. \end{aligned}$$

When one uses  $\eta = 0.1$  as the clustering radius, one obtains two clusters: one with average inverse weight 0.625 and the other with average inverse weight 1.25. If one then

uses eq. (32) to combine  $P$ -values, one attains the following results

$$\begin{aligned} \text{Prob}(Q \leq \tau) &= 1.472453 \times 10^{-6} + 1.171521 \times 10^{-7} + 0 \\ &\quad + 2.584710 \times 10^{-9} + 4.889899 \times 10^{-10} + \mathcal{O}(10^{-12}) \\ &= 1.59268 \times 10^{-6}, \end{aligned} \tag{34}$$

which contains no sign alternation and agrees well with the results from both (6) and (21). This illustrates the robustness of eq. (32) in combining  $P$ -values. Note that the third term on the right hand side of (34) is zero. This is because the multiplying factor  $Y_{k;3}$  is zero for both clusters. In general,  $Y_{k;3}$  measures the skewness of inverse weights associated with cluster  $k$  and for our case here both cluster of inverse weights are perfectly symmetrical with respect to their centers, leading to zero skewness. If the inverse weights of cluster  $k$  distribute perfectly symmetrically with respect to its center, it is evident from eq. (28) that  $Y_{k;g} = 0$  for odd  $g$ .

Evidently if one chooses a large clustering radius  $\eta$  and then uses eq. (32) to combine  $P$ -values, many higher order terms in the expansion will be required to achieve high accuracy in the final combined  $P$ -value.

## 7. Future directions

Although the expression (16) provides access to exact statistics for a broader domain of problems and our expansion formula (32) provides accurate and stable statistics even when nearly degenerate weights are present, there remain a few unanswered questions that should be addressed by the community in the near future. For example, even though we can accommodate any reasonable  $P$ -value weighting, thanks to (32), the more difficult question is how does one choose the right set of weights when combining statistical significance (Zelen and Joel, 1959; Koziol and Perlman, 1978; Hedges, 1985; Pepe and Fleming, 1989; Forrest, 2001). The weights chosen reflects how much does one wish to trust various obtained  $P$ -values. Ideally, a fully systematic method should also provide a metric for choosing appropriate weights. How to obtain the best set of weights remains an open problem and definitely deserves further investigations.

Another limitation of (16) and (32), and consequently of Fisher's and Good's formulas, is that one must assume the  $P$ -values to be combined as independent. In real applications, it is foreseeable that  $P$ -values reported by various methods may exhibit non-negligible correlations. How to obtain the correlation (Wei and Lachin, 1984; Pocock et al., 1987; James, 1991) and properly take into account the existence of  $P$ -value correlations (Brown, 1975; Kost and McDermott, 2002; Hou, 2005) is also a challenging problem that we hope to address in the near future.

## Acknowledgement

This work was supported by the Intramural Research Program of the National Library of Medicine at National Institutes of Health.

## References

Alves, G., W. W. Wu, G. Wang, R. F. Shen, and Y. K. Yu (2008, Aug). Enhancing peptide identification confidence by combining search methods. *J. Proteome Res.* 7, 3102–3113.

- Amari, S. and R. B. Mirsa (1997). Closed-form expression for distribution of the sum of independent exponential random variables. *IEEE Trans. Reliability* 46(4), 519–522.
- Bahrucha-Reid, A. (1960). *Elements of the Theory of Markov Processes and their Applications*. McGraw-Hill.
- Bailey, T. L. and M. Gribskov (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14, 48–54.
- Bateman, H. (1910). A solution of a system of differential equations occurring in the theory of radioactive transformation. *Pro. Cambridge Philosophical Soc.* 15, 423–427.
- Bhoj, D. S. (1992). On the distribution of the weighted combination of independent probabilities. *Statistics & Probability Letters* 15(1), 37 – 40.
- Brown, M. B. (1975). A method for combining non-independent, one-sided tests of significance. *Biometrics* 31(4), 987–992.
- Cox, D. (1962). *Renewal Theory*. Methuen & Co.
- Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Volume 2. John Wiley & Sons.
- Fisher, R. A. (1932). *Statistical Methods for Research Workers, vol. II*. Edinburgh: Oliver and Boyd.
- Forrest, W. F. (2001). Weighting improves the "new Haseman-Elston" method. *Hum Hered* 52(1), 47–54.
- Gabler, S. (1987). A quick and easy approximation to the distribution of a sum of weighted chi-square variables. *Statistische Hefte* 28, 317–25.
- Good, I. J. (1955). On the weighted combination of significance tests. *Journal of the Royal Statistical Society. Series B (Methodological)* 17(2), 264–265.
- Harrison, P. G. (1990). Laplace transform inversion and passage-time distribution in markov processes. *J. Appl. Prob.* 27(1), 74–87.
- Hedges, L. V., O. I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press.
- Hess, A. and H. Iyer (2007). Fisher’s combined p-value for detecting differentially expressed genes using Affymetrix expression arrays. *BMC Genomics* 8(NIL), 96.
- Hou, C.-D. (2005). A simple approximation for the distribution of the weighted combination of non-independent or independent probabilities. *Statistics & Probability Letters* 73(2), 179 – 187.
- James, S. (1991). Approximate multinormal probabilities applied to correlated multiple endpoints in clinical trials. *Stat Med* 10(7), 1123–35.
- Jasiulewicz, H. and W. Kordecki (2003). Convolutions of Erlang and of Pascal distributions with applications to reliability . *Demonstratio Mathematica* 36, 231–238.

- Kost, J. T. and M. P. McDermott (2002). Combining dependent  $p$ -values. *Statistics & Probability Letters* 60(2), 183–190.
- Koziol, J. and M. Perlman (1978). Combining independent chi-squared tests. *J. Amer. Statist. Assoc.* 73, 753–763.
- Koziol, J. A. (1996). A note on Lancaster's Procedure for the Combination of Independent Events. *Biometrical Journal* 38, 653–660.
- Koziol, J. A. and H. C. Tuckwell (1994). A weighted nonparametric procedure for the combination of independent events. *Biom. J.* 36, 1005–1012.
- Lancaster, H. D. (1961). The combination of probabilities: an application of orthogonal functions. *Austr. J. Statist.* 3, 20–33.
- Likes, J. (1967). Distributions of some statistics in samples from exponential and power-function populations. *Journal of the American Statistical Association* 62(317), pp. 259–271.
- Liptak, P. (1958). On the combination of independent tests. *Magyar Tud. Akad. Nat. Kutato int. Kozl.* 3, 171–197.
- Loesgen, S., A. Dempfle, A. Golla, and H. Bickeboller (2001). Weighting schemes in pooled linkage analysis. *Genet Epidemiol* 21 Suppl 1(NIL), S142–7.
- MacNeill, I. B. (1974). Tests for change of parameter at unknown times and distributions of some related functionals on brownian motion. *The Annals of Statistics* 2(5), pp. 950–962.
- Mathai, A. (1983). On linear combinations of independent exponential variables. *Communications in Statistics - Theory and Methods* 12, 625–632.
- McGill, W. J. and J. Gibbon (1965). The general-gamma distribution and reaction times. *Journal of Mathematical Psychology* 2(1), 1 – 18.
- Morse, P. (1958). *Queues, Inventories and Maintenance*. John Wiley & Sons.
- Mosteller, F. and R. R. Bush (1954). *Selected quantitative techniques*. In: *Handbook of Social Psychology*., Volume 1. Cambridge, Mass.: Addison-Wesley.
- Olkin, I. (1995). Statistical and theoretical considerations in meta-analysis. *Journal of Clinical Epidemiology* 48(1), 133 – 146. The Potsdam International Consultation on Meta-Analysis.
- Olkin, I. and H. Saner (2001). Approximations for trimmed Fisher procedures in research synthesis. *Stat Methods Med Res* 10(4), 267–76.
- Pearson, E. S. (1938). The probability integral transformation for testing goodness of fit and combining independent tests of significance. *Biometrika* 30, 134–148.
- Pearson, K. (1933). On a method of determining whether a sample of size  $n$  supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika* 25, 379–410.

- Pepe, M. S. and T. R. Fleming (1989). Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. *Biometrics* 45(2), 497–507.
- Pocock, S. J., N. L. Geller, and A. A. Tsiatis (1987). The analysis of multiple endpoints in clinical trials. *Biometrics* 43(3), pp. 487–498.
- Solomon, H. and M. A. Stephens (1977). Distribution of a sum of weighted chi-square variables. *Journal of the American Statistical Association* 72(360), pp. 881–885.
- Stouffer, S., E. Suchman, L. DeVinney, S. Star, and R. M. J. Williams (1949). *The American Soldier, Vol. 1: Adjustment during Army Life*. Princeton: Princeton University Press.
- Tippett, L. (1931). *The Methods of Statistics*. London: Williams and Norgate Ltd.
- Wei, L. J. and J. M. Lachin (1984). Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *Journal of the American Statistical Association* 79(387), pp. 653–661.
- Whitlock, M. C. (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher’s approach. *J Evol Biol* 18(5), 1368–73.
- Yu, Y.-K., E. M. Gertz, R. Agarwala, A. A. Schaffer, and S. F. Altschul (2006). Retrieval accuracy, statistical significance and compositional similarity in protein sequence database searches. *Nucleic Acids Res* 34(20), 5966–73.
- Zelen, M. and L. S. Joel (1959). The weighted compounding of two independent significance tests. *The Annals of Mathematical Statistics* 30(4), pp. 885–895.